Citation:

Villegas R, Gao YT, Yang G, Li HL, Elasy TA, Zheng W, Shu XO. Legume and soy food intake and the incidence of type 2 diabetes in the Shanghai Women's Health Study. *Am J Clin Nutr*. 2008 Jan; 87 (1): 162-167.

PubMed ID: 18175751

Study Design:

Prospective Cohort Study

Class:

B - Click here for explanation of classification scheme.

Research Design and Implementation Rating:

POSITIVE: See Research Design and Implementation Criteria Checklist below.

Research Purpose:

To examine the association between legume and soy food consumption and self-reported type 2 diabetes (T2D).

Inclusion Criteria:

- Participants in the Shanghai Women's Health Study (SWHS)
- Aged 40-70 years.

Exclusion Criteria:

- Women <40 and women >70 years.
- Those with T2D, cancer or cardiovascular disease (CVD)
- Those with extreme values for total energy intake (<500 or >3,500kcal per day).

Description of Study Protocol:

Recruitment

All eligible women aged 40-70 years who resided in seven urban communities in Shanghai, China were invited to participate.

Design

- Prospective cohort study
- Participants completed a detailed survey with a personal interview to assess dietary intake, physical activity, measurements of anthropometrics and other lifestyle factors at baseline (1997-2000) and biennally (2000-2002 and 2002-2004).

Dietary Intake/Dietary Assessment Methodology

- Dietary intake was assessed by a validated food-frequency questionnaire (FFQ) administered by trained personnel at baseline and again at the first follow-up survey
- If participants developed T2D, cancer or CVD only the baseline FFQ was used in the analysis
- If participants did not develop these diseases the mean of the baseline and follow-up was used
- The FFQ included 77 food items that represented 90% of foods commonly consumed in urban Shanghai during the study period.

Statistical Analysis

- The Cox proportional hazards model was used to assess the effect of food group consumption on the incidence of T2D
- Food groups (grams per day) were categorized by quintile distribution with the lowest quintile serving as the reference
- Tests for trend were performed by entering the categorical variables as continuous variables in the models
- All models were adjusted for the following potential confounding variables: Age, BMI, waist-to-hip ratio, total energy, energy-adjusted fiber intake, and vegetable intake (all entered as continuous variables) as well as income level, education level, occupation, physical activity, smoking status, alcohol consumption status, and presence of hypertension at baseline (as categorical variables).

Data Collection Summary:

Timing of Measurements

At baseline (1997-2000) and biennially (2000-2002 and 2002-2004).

Dependent Variables

- Total legumes and three mutually exclusive groups:
  - Soybeans (dried and fresh),
  - Peanuts, and
  - Other legumes
- Soy products such as soy milk, bean curd (tofu), fried bean curd, vegetarian chicken and bean curd cake
- Soy milk and “other soy products” were analyzed separately.

Independent Variables

Development of T2D (self-reported).

Control Variables

- Age
- BMI
- Waist-to-hip ratio (WHR)
- Total energy
- Energy-adjusted fiber intake

- Vegetable intake
- Income level
- Education level
- Occupation
- Physical activity
- Smoking status
- Alcohol consumption status
- Presence of HTN at baseline.

Description of Actual Data Sample:

- *Initial N:* 74,942 women in the SWHS
- *Attrition:* 64,191
- *Age:* 40-70 years
- *Ethnicity:* Chinese
- *Other relevant demographics:* None
- *Anthropometrics:* None
- *Location:* Shanghai, China.

Summary of Results:

The median intake of total legumes was 30.5g per day, for soybeans was 11.0g per day, for peanuts was 0.7g per day, and for other legumes was 15.5g per day. Total legume consumption and consumption of soybeans and other legumes were each associated with a decrease in risk of T2D:

- All legumes RR between extreme quintiles =0.62; 95% CI: 0.51, 0.74; P for trend <0.0001
- Soybeans RR between extreme quintiles =0.53; 95% CI: 0.45, 0.62; P for trend <0.0001
- Other legumes (not including peanuts) RR between extreme quintiles =0.76; 95% CI: 0.64, 0.90; P for trend <0.0001
- Soy milk RR between high and no intake = 0.61; 95% CI: 0.54, 0.70; P for trend <0.0001.

There was no significant association between consumption of other soy products or total soy protein and the risk of T2D.

Author Conclusion:

The authors concluded that consumption of legumes was inversely associated with the risk of T2D in this population.

Reviewer Comments:

Research Design and Implementation Criteria Checklist: Primary Research

Relevance Questions

1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (Not Applicable for some epidemiological studies)	Yes
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	Yes
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to nutrition or dietetics practice?	Yes
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	Yes

Validity Questions

1.	**Was the research question clearly stated?**	Yes
1.1.	Was (were) the specific intervention(s) or procedure(s) [independent variable(s)] identified?	Yes
1.2.	Was (were) the outcome(s) [dependent variable(s)] clearly indicated?	Yes
1.3.	Were the target population and setting specified?	Yes
2.	**Was the selection of study subjects/patients free from bias?**	Yes
2.1.	Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression, diagnostic or prognosis criteria), and with sufficient detail and without omitting criteria critical to the study?	Yes
2.2.	Were criteria applied equally to all study groups?	Yes
2.3.	Were health, demographics, and other characteristics of subjects described?	Yes
2.4.	Were the subjects/patients a representative sample of the relevant population?	Yes
3.	**Were study groups comparable?**	Yes
3.1.	Was the method of assigning subjects/patients to groups described and unbiased? (Method of randomization identified if RCT)	N/A
3.2.	Were distribution of disease status, prognostic factors, and other factors (e.g., demographics) similar across study groups at baseline?	N/A
3.3.	Were concurrent controls used? (Concurrent preferred over historical controls.)	Yes
3.4.	If cohort study or cross-sectional study, were groups comparable on important confounding factors and/or were preexisting differences accounted for by using appropriate adjustments in statistical analysis?	Yes

3.5.	If case control or cross-sectional study, were potential confounding factors comparable for cases and controls? (If case series or trial with subjects serving as own control, this criterion is not applicable. Criterion may not be applicable in some cross-sectional studies.)	Yes
3.6.	If diagnostic test, was there an independent blind comparison with an appropriate reference standard (e.g., "gold standard")?	N/A
4.	**Was method of handling withdrawals described?**	Yes
4.1.	Were follow-up methods described and the same for all groups?	Yes
4.2.	Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.)	Yes
4.3.	Were all enrolled subjects/patients (in the original sample) accounted for?	Yes
4.4.	Were reasons for withdrawals similar across groups?	Yes
4.5.	If diagnostic test, was decision to perform reference test not dependent on results of test under study?	N/A
5.	**Was blinding used to prevent introduction of bias?**	Yes
5.1.	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate?	N/A
5.2.	Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)	Yes
5.3.	In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?	N/A
5.4.	In case control study, was case definition explicit and case ascertainment not influenced by exposure status?	N/A
5.5.	In diagnostic study, were test results blinded to patient history and other test results?	N/A
6.	**Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were interveningfactors described?**	Yes
6.1.	In RCT or other intervention trial, were protocols described for all regimens studied?	N/A
6.2.	In observational study, were interventions, study settings, and clinicians/provider described?	Yes
6.3.	Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?	Yes
6.4.	Was the amount of exposure and, if relevant, subject/patient compliance measured?	Yes

6.5.	Were co-interventions (e.g., ancillary treatments, other therapies) described?	N/A
6.6.	Were extra or unplanned treatments described?	N/A
6.7.	Was the information for 6.4, 6.5, and 6.6 assessed the same way for all groups?	Yes
6.8.	In diagnostic study, were details of test administration and replication sufficient?	N/A
7.	**Were outcomes clearly defined and the measurements valid and reliable?**	Yes
7.1.	Were primary and secondary endpoints described and relevant to the question?	Yes
7.2.	Were nutrition measures appropriate to question and outcomes of concern?	Yes
7.3.	Was the period of follow-up long enough for important outcome(s) to occur?	Yes
7.4.	Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?	Yes
7.5.	Was the measurement of effect at an appropriate level of precision?	Yes
7.6.	Were other factors accounted for (measured) that could affect outcomes?	Yes
7.7.	Were the measurements conducted consistently across groups?	Yes
8.	**Was the statistical analysis appropriate for the study design and type of outcome indicators?**	Yes
8.1.	Were statistical analyses adequately described and the results reported appropriately?	Yes
8.2.	Were correct statistical tests used and assumptions of test not violated?	Yes
8.3.	Were statistics reported with levels of significance and/or confidence intervals?	Yes
8.4.	Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?	N/A
8.5.	Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?	Yes
8.6.	Was clinical significance as well as statistical significance reported?	Yes
8.7.	If negative findings, was a power calculation reported to address type 2 error?	N/A
9.	**Are conclusions supported by results with biases and limitations taken into consideration?**	Yes
9.1.	Is there a discussion of findings?	Yes

9.2.	Are biases and study limitations identified and discussed?	Yes
10.	**Is bias due to study's funding or sponsorship unlikely?**	Yes
10.1.	Were sources of funding and investigators' affiliations described?	Yes
10.2.	Was the study free from apparent conflict of interest?	Yes